FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

ParkDetect
Early Diagnosing Parkinson's Disease

Ricardo Graça

Mestrado Integrado em Engenharia Informática e Computação

Supervisor: João Pedro Carvalho Leal Mendes Moreira

July 19, 2013

ParkDetect
Early Diagnosing Parkinson's Disease

Ricardo Graça

Mestrado Integrado em Engenharia Informática e Computação

Approved in oral examination by the committee:

Chair: João Carlos Pascoal Faria (PhD)
External Examiner: Pedro Manuel Henriques da Cunha Abreu (PhD)
Supervisor: João Pedro Carvalho Leal Mendes Moreira (PhD)

July 19, 2013

Abstract

Parkinson's disease is one of the most common neurodegenerative disorders of the central nervous system that affects elderly. There are four main symptoms: tremors, rigidity, bradykinesia (slow movements) and posture instability. Nowadays any type of diagnose for this disorder is done through observation by a health care professional specialized in this area. Therefore it is necessary a method that is simple and efficient for health care professionals of general clinic to use so they can have some grounded backup to decide to forward a possible patient to a specialist. In this context a mobile application where a health care professional can insert values about possible patient's symptoms or the patient himself can realize small test is an interesting challenge due to reach they have today. This project can be split in three important phases: (1) development of a smartphone application, (2) use it to gather data from real patients and a control group and (3) testing and selection of a classification algorithm (selecting the relevant data and compare different algorithms) to be inserted in the same application. The first phase was the one with the most research since it was needed to understand how the symptoms could be detected only using the smartphone components and develop/adapt the application. The second one was the most time consuming, lasting from after the application getting developed to almost the end of the available time due to delays from the medical institution and the lack of capable patients that could perform the tests. The final one was highly affected with the lack of available data, making properly grounded conclusions impossible, however it was possible to obtain some promising results from the gait analysis of the patients where the pelvic sway was a good feature to help differentiate Parkinson patients from healthy ones.

Resumo

Parkinson é uma das doenças neurodegenerativas mais comuns do sistema nervoso central que afeta, principalmente, idosos. Existem quatro sintomas principais: tremores, rigidez, bradicinesia (movimentos lentos) e postura instável. Hoje em dia qualquer tipo de diagnóstico para esta doença é feito através das capacidades de observação de um profissional médico especializado na área da Neurologia. Assim é necessário um método simples e eficaz para profissionais médicos de clínica geral (médicos de família) que possam usar para tomar uma decisão mais fundamentada sobre o encaminhamento de um possível paciente para um especialista. Neste contexto uma aplicação para um dispositivo móvel é um desafio interessante devido à facilidade de acesso a um. Este projeto tem três importantes fases: (1) desenvolvimento de uma aplicação para um *smartphone*, (2) usa-la para recolher dados de pacientes de Parkinson e de um grupo de controlo e(3) por fim testar e selecionar diferentes algoritmos de decisão (passando por fases como seleção de dados relevantes). A primeira fase foi a que necessitou de mais pesquisa para se perceber como é que os sintomas da doença poderiam ser captados usando apenas componentes de um *smartphone* e desenvolver/adaptar a aplicação. A segunda fase foi a mais longa que durou desde a finalização da aplicação até, praticamente, o final do tempo disponível para a dissertação devido a atrasos por parte da instituição médica e a falta de pacientes capazes de efetuarem os testes. A fase final foi muito afetada com a baixa quantidade de dados disponíveis, impossibilitando a tiragem de conclusões bem fundamentadas, contudo foi possível obter resultados bastantes promissores na análise de locomoção onde a o balanceamento pélvico mostrou-se ser uma característica fulcral para diferenciar doentes de Parkinson de uma pessoa saudável.

Acknowledgements

In the first place I would like to thanks my family and girlfriend for all the support given, not only for the development of this dissertation but for the whole 5 years I spent in FEUP. Also appreciate all the help given by both my supervisors, Professor João Mendes Moreira and Msc Rui Castro. They were invaluable from the start helping with the implementation and all the necessary documentation and devices for this project. Thank all the support given by Fraunhofer, namely Silvia Rêgo for all the help gathering the subjects necessary for the control group and the support in S. João's Hospital. And finally thank Doctor João Massano, Doctor Carolina Garrett and Doctor Ana from Neurology Department of S. João's Hospital for their help, starting with the support with better understanding Parkinson's Disease to the gathering of Parkinson's Disease patients for this study.

Ricardo Graça

With Parkinson's, it's like you're in the middle of the street and you're stuck there in cement shoes and you know a bus is coming at you, but you don't know when. You think you can hear it rumbling, but you have a lot of time to think. And so you just don't live that moment of the bus hitting you until it happens. There's all kinds of room in that space.

Michael J. Fox, *Good Housekeeping*, June 2011

Contents

List of Figures

LIST OF FIGURES

List of Tables

Abbreviations

CA	Classification Algorithm
CFS	Correlation-based Feature Selection
FS	Feature Selection
H&Y	Hoehn and Yahr
IREP	Incremental Reduced Error Pruning
PD	Parkinson's Disease
ROC	Receiver Operating Characteristics
SA	Simulated Annealing
SBE	Sequential Backward Elimination
SFS	Sequential Forward Selection
UPDRS	Unified Parkinson's Disease Rating Scale

Chapter 1

Introduction

1.1 Context

Parkinson's disease (PD) is one of the most common neurodegenerative disorders of the central nervous system. This disorder is more common in persons with over 50 years of age and affects the lives of 2 million persons in Europe alone. This disease is named after James Parkinson, who published the first paper describing this problem in 1817. There are five symptoms that characterize this disease: tremor, rigidity, bradykinesia or slowness of movement, hand asymmetry and posture instability(Jankovic, 2008; Savitt *et al.*, 2006; Massano & Bhatia, 2012).

- **Tremor** — defined as an involuntary oscillating movement of one or more body parts of the patient that is only visible when the part in question is at rest, therefore is called rest tremor. This kind of tremor disappears when any voluntary movement is executed. This is the most common symptom and is, most of the times, one of the first to manifest its appearance.

- **Rigidity** — affects mainly the movement of joints caused by the excessive contraction of muscles.

- **Bradykinesia** — Slowness of movement affects the whole movement of the patient, although in early stages is more evident in the daily tasks such as writing, dressing or using tools.

- **Hand Asymmetry** — related with bradykinesia and the rest tremors. These symptoms normally only affects on side of the body resullting in an asymmetry between the movement of both hands.

- **Posture instability** — is only evident in later stages of PD.

Although there are more symptoms such as numbness, problems with speech, blurry vision, micrography (smaller handwriting), sleep disorders, muscle pain, they may not be truly related with PD.

1.2 Motivation

Nowadays the most accepted method used by specialists is the Unified Parkinson's Disease Rating Scale (UPDRS), however it is very experience dependent. It must be executed by a very experienced and skilled specialist that can observe different symptoms and rate them in a scale. Questions about daily living and some ordinary tasks are also made and affect the scaling. But, as previously mentioned, it cannot be executed by any person and a different tool would greatly help other health care professionals in this matter. Different studies have also shown progress in diagnosing PD analysing the speech or the spiral drawing of possible patients. Although both methods are in progress at this moment. At this moment the main problem in early diagnosing PD is that the early symptoms can easily be mistaken with old age ones and the patient is only forwarded to a specialist when the disease is already in a later stage making the treatment not so effective.

1.3 Research Question

The main question trying to be answered in this project is the following:

- Is there any possible way to a health care professional in general clinic to early diagnose PD so the patient can be properly forwarded to a specialist?

The main objective of this project is the development of an application to be used by a health care professional to verify the possible existence of PD so the patient can be properly treated by a specialist.

1.4 Research Goals

This dissertation proposes a mobile application for smartphones to be used by any health care professional. The project involved different phases:

1. Development of a mobile application for data gathering purposes.
2. Gathering real data from patients and healthy people.
3. Selection of the relevant data.
4. Creation of a machine learning classification model.
5. Development of a mobile application incorporating the model previously referred.

1.5 Document Structure

This document is divided in the following parts:

- **Chapter 3: State of the Art** — part where the advantages and disadvantages of the currently available options to diagnose PD are presented and discussed, as well the different methods that can be adopted to solve this kind of classification problem.
- **Chapter 4: Implementation** — part where the implementation of the different components of the application are explained and as well the features that can be extracted by each. Also how the different machine learning algorithms were tested.
- **Chapter 5: Results** — part where all the results obtained from all the different components are displayed. Necessary results to validate some decisions made during the implementation phase are also shown. The discussion about the results,pointing their importance and relevance for this study are also in this chapter.
- **Chapter 6: Conclusions and Future Work** — final part where it is made a small retrospective of all the project is made and some new components that should be added to this project are named and explained.

Chapter 2

State of the Art

This project convolved two kinds of state of the art. One about the problem of diagnosing neurodegenerative diseases, in particular PD, where the current method is presented but also new innovative methods are simply explained. The second one is about the methods used to solve this problem in a more technological way, where different methods are theoretically compared and among a long list of possibilities, a final way is narrowed to be explored in the later stages of the project.

2.1 Problem

PD is most common in elderly. Aging produces several modifications in the human brain like shrinking (Peters, 2002). This means there is a decline in brain performance (learning, memory and reflexes) due to deterioration of synaptic contact and in neurotransmitters and neurohormones (Nieto-Sampedro & Nieto-Diaz, 2005). One way to delay the appearance of neurodegenerative diseases like Alzheimer's or Parkinson's is submit to a diet rich with anti-oxidant and anti-inflammatory agents (curcumin, green tea and feluric acid)(Farooqui & Farooqui, 2009).

Nowadays there are several diagnostic tools to discover neurodegenerative diseases like PD, however they can be expensive and not fully reliable. The tools available to diagnose PD depend on the experience of the specialist, however some automatic support decision tools are on developing or testing phases.

2.1.1 Neurodegenerative diseases diagnose

Several neurodegenerative diseases have been linked to oxidative damage caused by oxidative stress. Oxidative stress refers to cytotoxic consequences caused by the process of the use of molecular oxygen by a cell. This damage is also inflicted by Reactive Oxygen Species (ROS) when presented in high levels. At low levels it poses a main role in cell activities like growth and adaptation responses. ROS can also be caused by stress (Allen & Tresini, 2000), or through

the interaction of polyunsaturated fatty acids that also leave traces of metal ion $Fe^{+}3$ (Qin *et al.*, 2002).

Lipidomics and proteomics have emerged as important technologies (German *et al.*, 2007; Watson, 2006; Bowers-Gentry *et al.*, 2006) for the identification of oxidative stress markers which has facilitated their connection with neurodegenerative diseases (Lu *et al.*, 2006). After several tests analysing cerebrospinal fluid of different neurodegenerative patients there is a pattern showing that there are increased levels of α-synuclein (a biomarker) in PD patients. In combination with brain-imaging techniques (SPECT, PET and MRI) there is a significant improvement in the detection of neurodegenerative diseases. Early diagnosis are the key for efficient long term treatment however the combination of all these techniques are only in range of the richest parts of the society.

2.1.2 Parkinson's Disease Diagnose

2.1.2.1 Traditional Diagnose

Most specialists diagnose PD using UPDRS which is a clinical metric to quantify PD impairment. It consists on 42 items where some may have some subdivisions and we shall refer to each item and subdivision as a section. There are four main major components UPDRS relies on:

1. Behaviour, mood and psychological state (4 sections, 1-4);
2. Daily routine activities (13 sections, 5-17);
3. Motor, regarding muscle control (27 sections, 18-44);
4. Therapy related complications (11 sections, 45-55).

The third component, many times designed as motor-UPDRS, is the one that most influences the UPDRS score and the fourth one is not used on non treated patients. The motor-UPDRS scores between 0 and 108, where 0 represents a healthy patient and 108 represents one with severe disabilities. The total UPDRS score is calculated summing up the score in each component ranging between 0 and 176 in not medicated patients (first three components) (Fahn & Elton, 1986).

Another vastly used method is the Hoehn and Yahr (H&Y) that provides overall PD stage assessment. This scale goes from 1 to 5 (having 1.5 and 2.5 in the modified version) where at 1 the patient shows unilateral involvement only (tremors in one hand for example) and at 5 means that the patient is confined to a bed or wheelchair. Studies have shown that this method can be mapped to the UPDRS (Tsanas *et al.*, 2012a).

However both methods require high experience and skill of the specialist to assert some physical and psychological conditions of the patient.

2.1.2.2 Using Speech

Recent studies have shown that speech may be used to help discriminate PD patients from healthy ones (Little *et al.*, 2009; Sapir *et al.*, 2010) which indicates that most PD patients may have a

vocal disorder (Ho *et al.*, 1998). Research already proved that a vocal disorder may be one of the first symptoms to appear nearly 5 year before clinical diagnose (Harel *et al.*, 2004). The main symptoms noticed on speech include loudness, vocal tremor and breathiness. Vocal disorders important for PD are dysphonia (inability to produce normal vocal sounds) and dysarthria (difficulty in pronouncing words) (Baken & Orlikoff, 2000). The severity of a vocal disorder is often evaluated using sustained vowel phonations or running speech, although the former is more complex to analyze due to linguistic confounds (Titze, 2000; Schoentgen & Gucteneere, 1995). Research has shown that the use of the sustained vowel "ahhh" is sufficient for the assessment of many voice disorders including the ones affecting PD patients, either to diagnose (Little *et al.*, 2009) or monitor (Tsanas *et al.*, 2010; Tsanas *et al.*, 2011). In (Little *et al.*, 2009) it is shown that voice processing algorithms could discriminate PD patients from an healthy ones with approximately 90% classification accuracy using four dysphonia features. A more recent study has increased the number of features used increasing its precision to 93%, although through the use of a group of Feature Selection Algorithms they were able to narrow them to half and get 99% precision (Tsanas *et al.*, 2012b). The main downside of this approach is that it was used a dataset captured in a studio. Consequently, the samples had high quality affecting the good quality of the final results. Taking this fact in consideration a more recent study had as main objective the assessment of a cellular mobile telephone network to the same purpose (Tsanas *et al.*, n.d.). Although some clinical relevant information can be retrieved, having a small margin of error comparing to the UPDRS (4.1 points for males and 3.8 points for females), this method used a simulated mobile network not taking in account variables such as noise, cellphone microphone noise or network failures.

2.1.2.3 Spiral Analysis

This method tries to diagnose PD mainly focusing on its most common symptom, rest tremor. In this project the subject needs to draw a spiral so that its shape and consistency can be evaluated. This project uses a touch screen tablet so the subject can execute the test everywhere and this way it does not have to be performed in a hospital where possible stress due to the environment can influence the result. Although this software does not replace a specialist it might give relevant information about the patient, such as the stage of the disorder. The Figure 2.1 shows a preliminary analysis of two subjects of testing where the left side represents a spiral and respective polar coordinates drawn by an healthy person and the right side represents a patient with PD. It still is in early development and it seems to be very promising, being the future work planned by the authors the development of a classification model to validate the usefulness of the application (Surangsrirat & Thanawattano, 2012).

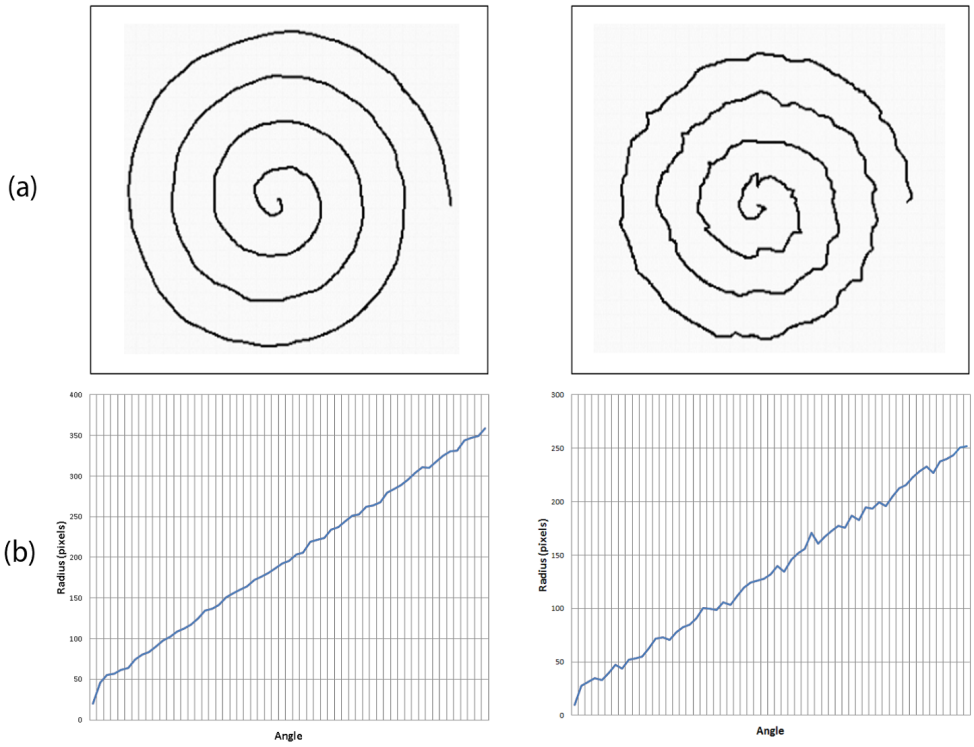

Figure 2.1: Spiral Comparison Healthy vs PD (Surangsrirat & Thanawattano, 2012)

Chapter 3

Methodologies

The solution for this problem will be a smartphone application. Since it is supposed to reach as much users as possible it needs to be implemented in a platform that has high levels of popularity and usage. Another important characteristic is that it needs to be easily deployed. Also it must include easy access to the smartphone components like the accelerometer and the gyroscope. The main candidate that has all the necessities is a smartphone with the Android Operating System. According to Gartner Inc. Android users are increasing at the highest rate comparing with other mobile operating systems and will become the leader in mobile platforms (Gandhewar & Sheikh, 2011). This application will be developed according to a Feature Selection phase where it will be decided which data of symptoms will be most relevant to the classification algorithm. In an early phase a data capturing application will be developed to extract values from different symptoms. It will use features like scales, direct interaction with the patient (spiral test and crash course test, the latter will demand the patient to navigate through a test course with the smartphone attached to his hips and values like time, pauses, difference from the stipulated path will be acquired) and yes/no questions. It is important to refer that all the data used will be acquired from real patients with the help of the Dr. João Massano. A control group will also be used for classification purposes.

3.1 Feature Selection

Feature Selection (FS) is an important step for any problem involving machine learning. FS is generally used to handle a full dataset and turn it into a subset with the best balance between size and relevance of data selecting the most relevant features. However the concept of relevance differs depending on the final objective. FS algorithms can be classified in two ways: the ones that manipulate the dataset giving a (weighed) order of features and those that cut off irrelevant features leaving a subset of original values. The concept of relevance of a feature depend on a variety of factors, among them, probably the most important, is the relevance of a feature to the final objective (two instances can only be classified thanks to the feature in question) (Molina

et al., 2002).
A FS algorithm is directly connected to a Classification algorithm (CA) and this relationship can be shown in three different schemes (all represented in Figure 3.1):

(a) **Embedded** — the classification algorithm has its own FS algorithm. Algorithms like decision trees or neural networks use this kind of scheme (Mitchell, 1982);

(b) **Filter Scheme** — this type of FS algorithm is independent from the classification algorithm where the former acts as a filter to irrelevant features before the learning phase. There are two types of algorithms for this scheme:

- Ranking — where the features are ranked based on their individuals evaluations;
- Subset — where a subset of features is selected evaluating the worth of it by considering the individual predictive ability of each feature along with the degree of redundancy between them

(c) **Wrapper Scheme** — In this type the FS algorithm acts in a opposite way comparing to the filter scheme using the CA results with different subsets at each iteration (John *et al.*, 1994). The main disadvantage of this method is the computational burden that comes from calling the classification algorithm several times.

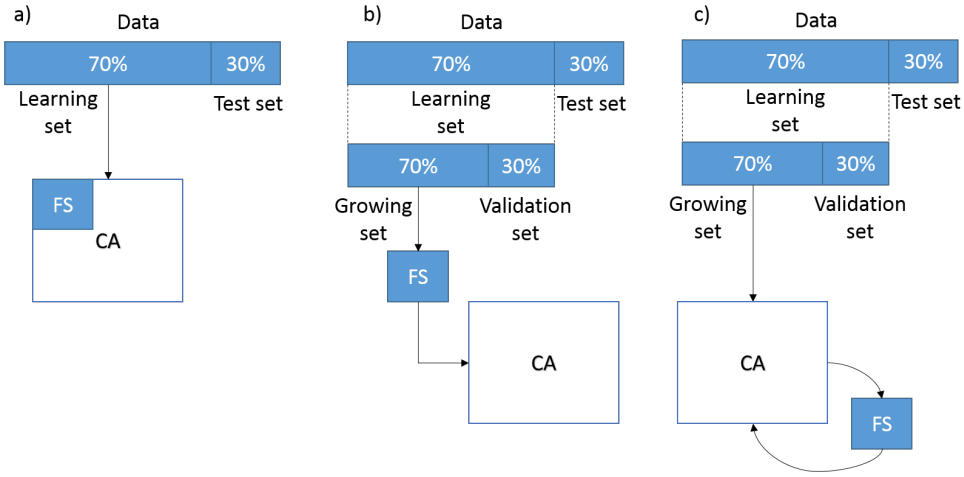

Figure 3.1: Feature Selection Schemes - a) Embedded, b) Filter and c) Wrapper

There are several feature selection algorithms which can be used for one problem and a brief description of the main ones are in the following sections.

3.1.1 FS Algorithms

- Chi (χ^2 Statistic)
  Chi-Squared is the most common statistical test that allows to check the difference between

the distribution expected and the one that the user has, if the feature is actually independent from the class value. Using a Chi-Squared it is possible to detect if a given distribution of values of a feature and classes are independent, that is, they are related and the feature is relevant to the class.

- Euclidean Distance
  Is the most common use of distance in many areas. In FS it is calculated the distance between each feature to all other features to check for similarities between features. If a distance is big it means that it has relevance for the class.

- Information Gain
  This method measures the number of bits necessary to encode the information (in this case the distribution of instances among classes). If the number of bits necessary to encode the information is high it means it has high entropy. Therefore a function measuring entropy must increase when the class distribution gets more spread out and be able to be applied recursively to allow finding the entropy of subsets of the data.

- Correlation-based Feature Selection (CFS)
  CFS searches feature subsets based on the degree of redundancy in the features. It aims to find subsets with features that, individually, are highly correlated with the class but have low inter-correlation. To overcome this downside of univariate filters, the multivariate filters were created that evaluate the worth of a subset taking in account the individual predictive ability of each feature and the degree of redundancy between them. The relevance of a subset grows with the correlation of features and classes and lowers with growing inter-correlation. CFS is used to determine the best feature subset and is normally combined with strategies like forward selection, backward elimination, best-first search and genetic search.

- Sequential Forward Selection (SFS)
  SFS is the simplest greedy search algorithm. It starts from a empty set and, sequentially, adds a feature that becomes the highest objective function when combined with the rest of the already selected features. SFS works best when the optimal subset of features is small. However after adding a feature, all other previously added cannot be removed even if they become obsolete.

- Sequential Backward Elimination (SBE)
  SBE works on opposite of SFS since it starts with a full set of features and sequentially remove the feature with the smallest decrease on a objective function. SBE works best on when the optimal feature subset is large. The main disadvantage is that an already discarded feature cannot be reevaluated.

- Simulated Annealing (SA)
  SA is often used on optimization problems. To form a strong alloy it is necessary to heat up the components, mix them and let the material cool to a determined temperature to avoid

defects. Comparing to a classification problem the initial state is a guessed solution and it is heated to a starting temperature. While at that temperature the system is allowed to change states until a equilibrium is reached (equilibrium means a low cost function). When equilibrium is reached the temperature decreases and yet again the system is allowed to change states. This process is repeated until equilibrium has been reached at a low temperature (which is called frozen) and a final solution has been obtained.

- Randomized Hill-Climbing
  Hill-Climbing is probably the most known algorithm for local search. It starts at a randomly generated state and moves to the neighbour with the best evaluation value. If a strict local-minimum is reached then it restarts at other generated state. This procedure is repeated until the solution is found. Since it checks all neighbours before moving to one of them this algorithm can take a lot of time.

This way the FS algorithm will depend on which Classification Algorithm is to be used since some of them might already have on embedded.

3.2 Selection of the Classification Algorithm

The selection of a classification algorithm depends on different factors, however the first thing we need to be aware is that there is no best algorithm for any given problem. The main factors that need to be considered are: Precision, velocity and scalability, robustness and interpretability.

3.2.1 Precision

One of the characteristics of this project is that it will have a relatively small dataset that narrows the possibilities to proper evaluate its precision. Due to this fact using Cross-Validation is the most common method that allow us to compare models with different parameters. In this method the data is split into K subsets of the same size where a different one is selected for testing and the remainder for training in each iteration (total of K times) as seen in Figure 3.2. It is necessary to be careful with the complexity of the created model since it may have high precision due to its high complexity (that can be related with data over fitting) (Hastie *et al.*, 2001).

Yet another method to evaluate the precision of a classification algorithm is a Confusion Matrix. It keeps count of every instance classified (correctly or not, see Figure 3.3). With this method it is easier to detect which errors are most common. Taking into consideration that this is a medical diagnosis problem the consequences of a miss-classification can be severe. This way a cost matrix can be defined to associate a cost to each type of classification error so instead of using the precision formula to evaluate it, it is used an average cost (Nisbet *et al.*, 2009).
Another method to determine not only the precision of an algorithm but also its accuracy is the Receiver Operating Characteristics (ROC) curve (example in Figure 3.4, where it represents the diagnosis of Tuberculosis (TB) through the value of pleural effusions and being Cancer in the negative cases). It has become quite useful in medical decision making. The ROC curve uses a test

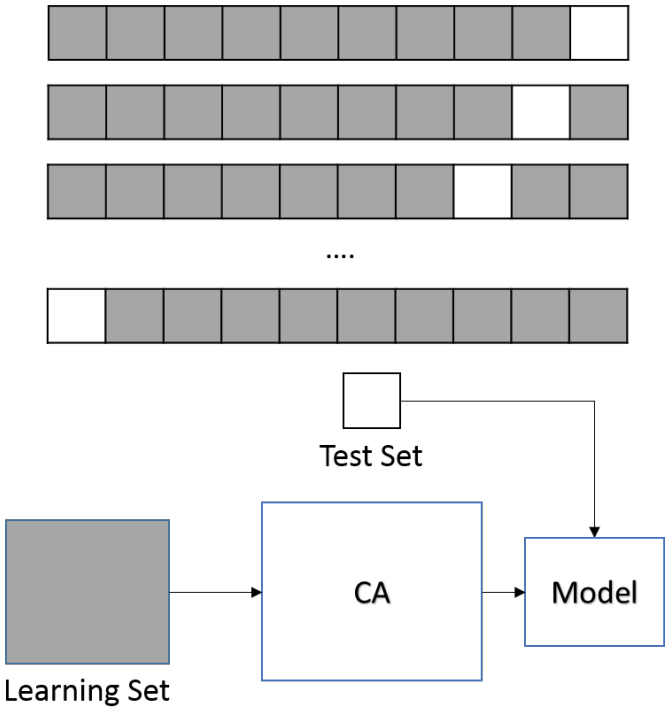

Figure 3.2: Cross Validation Example

		Actual Class	
		True	False
Predicted Class	True	True Positive	False Positive
	False	False Negative	True Negative

Figure 3.3: Confusion Matrix Example

set of data to be applied on the model where every time a true positive is classified the curve goes up but if it is a false positive it goes right. With this in mind it is easy to see that a ROC curve that increases very fast upwards will have a good classification rate. But comparing algorithms only by comparing curves by hand is not feasible so there is a value that can be extracted from the ROC curves: the Area Under Curve (AUC). Just like the name suggests the area under the curve can be calculated (taking in account that the axis go from 0 to 1, the area can only vary between the same values) where the algorithm with higher AUC is the better one (Hopley & van Schalkwyk, 2001).

3.2.2 Velocity and Scalability

There are two phases to assert this factors:

- Build the Model
- Use the Model

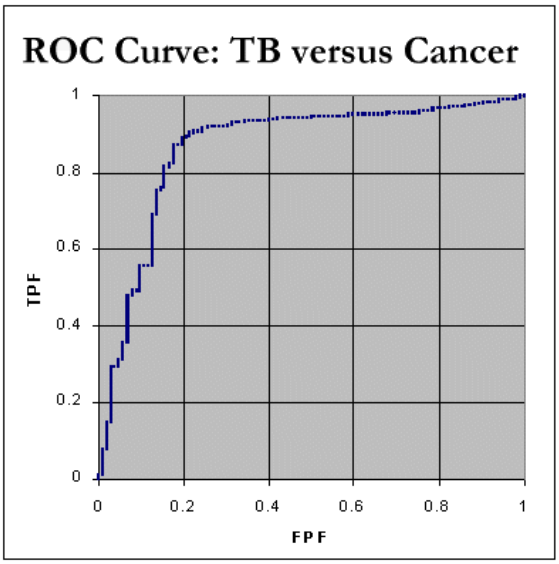

Figure 3.4: ROC Curve - TB vs Cancer (Hopley & van Schalkwyk, 2001)

This is very important in big datasets where going through them several times to build or use the model can be unthinkable because of the necessary memory and computation, however the dataset to be used is going to be small and these two factors do not impose an important role to the final decision.

3.2.3 Robustness

There are several situations that need to be taken into consideration when using the model:

- **Noise** — Miss introduced values (e.g. 650 instead of 65 in an age value). In this problem giving that all the data introduced is validated by the application (all range of values are limited) it is not expected any of these situations;

- **Missing Values** — In this case, due to the fact of inexperience of the doctor or inability of the patient to perform any test (physical limitations) some values might be left blank and it's important for the model to be able to deal with it or a default value must be used;

- **Irrelevant Characteristics** — Given that this project will go through a Feature Selection phase there will be no irrelevant features inserted by the user.

3.2.4 Interpretability

Some algorithms allow the user to verify how a classification was achieved and in this area of business (health care) is extremely important. A doctor must be able to check which symptom most affect the decision and for this reason this factor is the most relevant one.

3.3 Classification Algorithm

Being the interpretability the most important factor, it was possible to narrow the possibilities to three kinds of algorithms: Decision Trees, Classification Rules and Bayesian Networks. The other factors will be analysed during the implementation phase since that was only possible to achieve after having data from some test subjects.

3.3.1 Decision Trees

Most of the research in this field was done by Ross Quinlan (Quinlan, 2008) and at the moment one of the most used and with better characteristics is the C5.0 (RuleQuest, 2012). However its predecessor (the C4.5) is highly capable to deal with problems with low amounts of data since the improvements applied were on the scalability and on the feature selection of problems with an elevated number of features. This type of model consists on a tree structure similar to a flowchart where each node represent a test to an attribute, each branch a possible result to the test and the leafs another test or a classification class. In the latter case the node is named a leaf node (Figure 3.5 ith a simple example of getting up early in the morning decision).

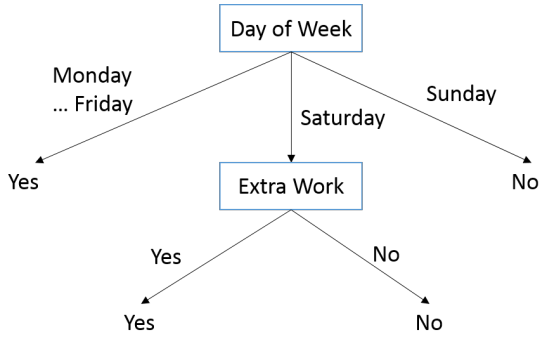

Figure 3.5: Decision Tree example (Getting Up Decision)

Building a decision tree has two phases:

- **Building the tree** — The main objective is having the most relevant and important features on the top of the tree. Through the use of heuristics or statistical measures it is possible to detect which feature amongst the dataset has the most information gain for example. That feature is selected and a test is associated to that node with that feature (a value under or above a detected point of difference in case of numerical values). This process is repeated until all branches end in a leaf node.
- **Pruning** — Identify and remove branches that represent noise or that are over fitting the dataset.

In the classification phase the new instance's features pass through all the test to get a final classification to the instance.

3.3.2 Classification Rules

Classification rules are relatively easy for people to understand (Cattlet, 1991) (see Figure 3.6 relating to the decision tree in Figure 3.5) and they outperform decision trees on a variety of problems (Pagallo *et al.*, 1990; Quinlan, 1987; Weiss & Indurkhya, 1991). At a first point they were fairly poor dealing with large noisy datasets however due to the nature of this project this would not be a problem. There are two main algorithms to be explored the RIPPERk (Cohen, 1995) and C5.0rules (RuleQuest, 2012)

```
IF day_of_week=Monday OR day_of_week=Tuesday OR ... Friday
THEN getUp=Yes

IF day_of_week=Sunday
THEN getUp=No

IF day_of_week=Saturday AND Extra_work=Yes
THEN getUp=Yes

IF day_of_week=Saturday AND Extra_work=No
THEN getUp=No
```

Figure 3.6: Classification Rules example (Getting Up Decision)

RIPPERk This algorithm is an evolution of the Incremental Reduced Error Pruning (IREP) (Fürnkranz & Widmer, 1994). The IREP integrates the Reduced Error Pruning (REP), used in Decision Trees as well, with a separate-and-conquer rule learning algorithm. IREP creates rules in a greedy way, one rule at the time and deletes every instance covered by it from the data (both positive and negative). This process is repeated until no more data is present or when an unacceptable error rate is reached. With some optimization to the stop condition and rule-value metric this algorithm was improved but the main component that created the RIPPERk was the ability to optimize even further the rules. The latter was achieved by creating two new alternative rules for each rule learned where greedily new conditions were added to these rules and the final rule (between the three) is selected using the MDL Heuristic. This process can be done k times improving the optimization even further at the cost of more time and space that scales nearly linearly (Cohen, 1995).

C5.0Rules This algorithm is based on C5.0 Decision Trees (RuleQuest, 2012) and suffers from the same problems as the trees where it may over fit the data however through the process of pruning this can be corrected. In comparison with RIPPERk this algorithm performs in a similar level in relatively small datasets but in large datasets it starts to fall behind because it has higher scaling rates than the opponent (Cohen, 1995).

3.3.3 Bayesian Networks

Bayesian Networks (BN) is another Classification Algorithm that has high interpretability. The induction is done through the use of probabilities (use of the Bayes' Rule in Equation 3.1)

$$P(a \mid b) = P(a)\frac{P(b \mid a)}{P(b)} \tag{3.1}$$

- $P(a \mid b)$ — Probability of a happening given that b happens.
- $P(a)$ — Probability of a happening.
- $P(b \mid a)$ — Probability of b happening given that a happens.
- $P(b)$ — Probability of b happening.

A BN consists on a set of variables and a set of directed edges between variables. Each variable has a finite set of mutually exclusive states (independence of states), all states and edges must form a directed acyclic graph and each edge between two variables (a -> b) represent a probability (P(b | a)) (see Figure 3.7 for a simple example (Pearl, 2003)). Another characteristic of BN is that it allows a consistent combination of information from various sources. The main disadvantage of BN is that it is NP-hard (non-deterministic polynomial-time hard), meaning that in some cases it may be costly to use it (Charniak, 1991).

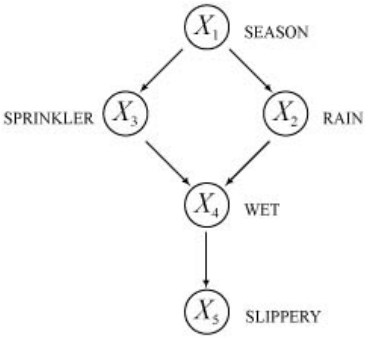

Figure 3.7: Bayesian Network Example (Pearl, 2003)

3.4 Final Remarks

Exploring the different options the final decision falls to the precision values of the different algorithms given the data to be learned, however this data will be collected during all the development of this project and a decision must be taken with only a small part of it. But after having some of the data (control group and patients) it will be possible to finally choose the "best" algorithm to this situation.

Chapter 4

Implementation

The implementation for this project consists in three steps:

1. Data Gathering Application
2. Data Analysis and Algorithm Implementation
3. Readjustment of the application (adding the classification model)

The first application developed was intended to gather all information possible with a set of components decided by throught the analysis of the state of the art and in some meetings with Dr. João Massano from the S.João's Hospital. The components are: spiral analysis, simple questions, gait analysis and tap analysis and will be thoroughly described in the following sections. Having a considerable number of instances all the data will be processed and analyzed using a FS algorithm (if necessary) and different classification algorithms (described in **??**). Having the results analyzed the model achieved will be inserted into the application and all the non relevant and unnecessary features will be removed.

4.1 Data Gathering Application

This application is a combination of different components that as a whole can gather physical data of the user both in a manual and in a automatic way. The components can involve direct and indirect interaction from the patient and basic observation skills from the health care professional.

4.1.1 Spiral Analysis

From the work explored in the State of the Art phase it stood obvious that PD diagnosis through the analysis of a spiral drawn by a person is effective and accurate. With this approach it is possible to evaluate movement disorders such as tremor, rigidity and dradykinesia (Surangsrirat & Thanawattano, 2012; Pullman, 1998). In the past few years a lot of work has been applied

in this area to prove the concept (Pullman *et al.*, 2008; Liu *et al.*, 2005; Miralles *et al.*, 2006; Aly *et al.*, 2007) and to test several implementations in different platforms using different kinds of spiral (pentagon, octogonal, Archimedian) (Surangsrirat & Thanawattano, 2012; Westin *et al.*, 2010; Wang *et al.*, 2008; Wang *et al.*, 2011; Cunnungham *et al.*, 2009).

The main concern of using it as a component of the application is the screen size of the device. To solve this problem a two sided approach was adopted. First the use of a smartphone with a considerable big screen (4 inches), second the use of a stylus that allows the user to see what he is drawing. The main problem of a stylus for modern touch surfaces is that it needs to be pressed by something equivalent to a finger's end size, but covering all the area the user wants to click. The solution was created by Adonit (Adonit, 2013) that has a stylus with a transparent disk on its end so the user can see exactly what he is drawing (see Figure 4.1).

Figure 4.1: Adonit Jot Classic Stylus (Adonit, 2013).

Having dealt with the main issues this component was inserted as a feature gathering component of the application to validate its usage. In it the patient tries to draw an Archimedean Spiral as perfect as possible (a water marked one is shown so the user tries to replicate it).

First thing that needed to be explored was how the coordinates of an Archimedean Spiral were calculated. A spiral of this kind has its points in a Polar Coordinate System and the smartphones screen uses a modified Cartesian Coordinate System (see Figure 4.2).

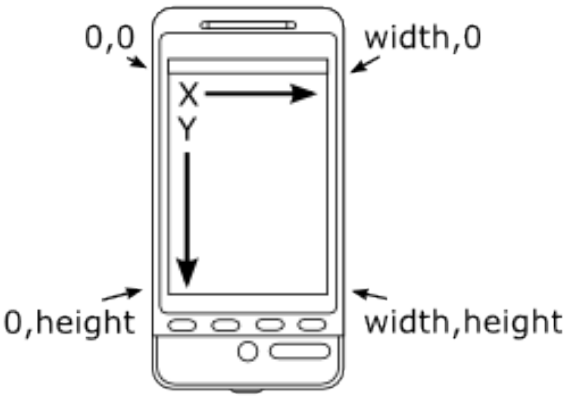

Figure 4.2: Android's Screen Coordinate System (TekEye, 2013)

In order to convert these coordinates from one to another system the center of the spiral on the screen (x_0,y_0) was defined and the values of (x,y)(Cartesian coordinates) and (r, θ) (spiral radius and angle of each point) were calculated through the Formulas 4.1, 4.2, 4.3 and 4.4:

$$x = r \times \sin(\theta) + x_0 \tag{4.1}$$

$$y = r \times \cos(\theta) + y_0 \tag{4.2}$$

$$r = \sqrt{(x - x_0)^2 + (y - y_0)^2} \tag{4.3}$$

$$\theta = \arctan(\frac{y - y_0}{x - x_0}) \tag{4.4}$$

Transforming the coordinates of the spiral drawn by the user to the Polar Coordinate System it is possible to evaluate them using the Archimedean Spiral's formula (see Formula 4.5),where r is the radius, θ the angle, a the spiral's orientation and b the distance between loops.

$$R = a + b \times \theta \tag{4.5}$$

Iterating through all points of the spiral it is possible to obtain the following features:

- **Average Error($\bar{\varepsilon}$)** $= \frac{\sum_{i=1}^{N}(|r - r_i|)}{N}$, between the drawn spiral and the perfect one comparing all pixels (N) of the patient's spiral (r_i) with the perfect radius for that angle(r).
- **Maximum Error** $= \max(\mid R - r_i \mid)$, going through all the pixels drawn(r_i) and comparing with the perfect ones(r) it is possible to detect the maximum error performed.
- **Standard Deviation Error** $= \sqrt{\frac{\sum_{i=1}^{N}(r_i - \bar{\varepsilon})}{N}}$, statistical measure to determine the distribution of the error in the drawn spiral.
- **Cross Ratio** $= \frac{crosses}{N}$, percentage of times that the user crosses the perfect spiral. Going through all the pixels (N) drawn the number of times that the error changes from negative to positive of vice-versa is counted (*crosses*).
- **Pressure Ratio** $= \frac{\frac{\sum_{i=1}^{N}(pR_i)}{N}}{\frac{\sum_{i=0}^{N}(pL_i)}{N}}$, while the spiral is being drawn a value of pressure on the screen is captured. A study has shown that PD patients often apply less pressure to one hemispiral

than to the other (Yu *et al.*, 1997) (pR_i - pressure applied on the right hemispiral and pL_i - pressure applied on the left hemispiral).

- **Number of Points Ratio** $= \frac{N_R}{N_L}$, depending on the drawing speed of the user in each hemispiral there may be a different number of pixels drawn in each side. Therefore a ratio between both sides (N_R - right side, N_L - left side) may indicate an asymmetry.

- **Time** $= t_N - t_1$, total time took to draw the spiral calculating the difference in time between the last point (t_N) and the first one (t_1).

4.1.2 Simple Questions

This project is focused on automatic assessment of symptoms of PD but information coming from simple observation where no specialized skill or experience is necessary may also prove relevant and useful. For PD two of the main symptoms that are often the first ones to manifest and endure during all the disease are: rest tremor and flexed posture. With a simple yes or no answer the observer (the healthcare professional) determines if the patient has any of these easily noticeable symptoms and inputs the answer to the application. Another goal with these two features is to see if they can be noticeable in the automatic assessment of the rest of the application.

4.1.3 Tap Analysis

As previously stated, this project had a protocol established with Hospital S.João. From that institution, Dr. João Massano helped with some very important notions about PD. In one of the meetings we discussed the symptoms that reveal themselves in an early stage of PD and how they can be diagnosed. Other symptoms mentioned, apart from rest tremor and flexed posture, were bradykinesia and hand asymmetry. The method discussed to diagnose them us by performing a series of repetitive movements. For this task and to keep the patients as concentrated as possible but also keeping them relaxed, a project already developed in Fraunhofer AICOS was adapted. It consists on a set of simple games that require the user to be as fast as possible with the use of smartphone's screen. For the dissertation project two of the games were integrated with the main application. The first one consists on gathering response time and pressure on the screen with each hand playing a game where the user must click on moles whenever they appear with his index finger (see Figure 4.3). To standardize this process the moles only appear on the points shown on the Figure 4.4.

The second one gathers click velocity, average time between clicks and pressure applied on the screen by playing a different game where the user must fill buckets of water as fast as possible by clicking a button several times with his thumb while holding the phone on his hand (see Figure 4.5).

From both there games it is possible to compare the performance of both hands comparing them to confirm the existence of a slightly higher difference (since it is expected that a small difference exists in every person due to being right/left handed having better reflexes in that one).

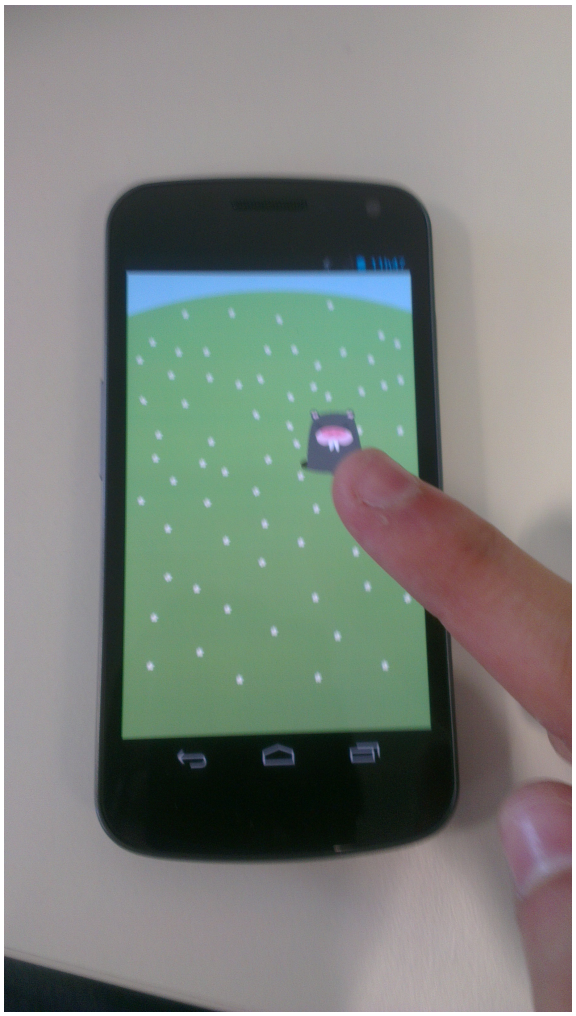

Figure 4.3: Layout of the first game incorporated into the project. Also it shows how the test is performed.

However some modifications regarding the logging procedures of these applications were necessary due to the fact that no information about timestamps of the taps or pressure applied were not being stored/calculated. The features gathered are:

- **Tap Time** — Response time (from when the mole appears until it is clicked). The values calculated are: average, standard deviation, maximum and minimum for each hand. Also the ratio between the averages is calculated $\frac{\text{right hand average}}{\text{left hand average}}$.
- **Tap Down Time** — Time spent with the finger pressed on the screen while clocking on a mole. Like the previous feature the same values are calculated, including the ratio between the averages.
- **Tap Pressure** — Pressure applied on the screen on each click. However it does not follow a regular pressure formula. This value stands between 0 and 1 (being 1 the logical maximum

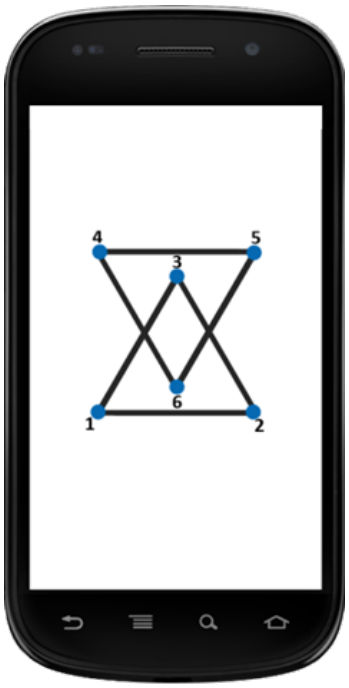

Figure 4.4: Possible positions where a mole can appear on the first game.

pressure but this value can be over 1 according to the Android Software Development Kit). This pressure is calculated according to the area of the finger being used to press the screen, the larger the area, the higher the pressure obtained. Following the previous features the same statistical measures and ratio are obtained.

- **Water Tap Time** — From the second game this value represents the time between each click filling the bucket. The same statistical measures are calculated.
- **Water Tap Pressure** — The same method from the Tap Pressure is used here and the same values are calculated.
- **Click Velocity** — This value represents the number of clicks per second a person can do. In this case the value is calculated using $\frac{\text{number of clicks}}{\text{total time between clicks}}$. In this case only the value of the speed is calculated for each hand and the ratio between them.

4.1.4 Gait Analysis

Another way of diagnosing bradykinesia is through the analysis of the patient's gait. This is possible since one of the symptoms of PD is slowness of movements that causes, for example, shorter steps. Using a project developed in Fraunhofer AICOS (Guimarães, 2011), the patient only needs to perform a simple walking test with a fixed duration (25 seconds) in a straight line. This test allows to read data from the accelerometer and gyroscope of the device. The smartphone is positioned on the back of the trunk of the user using a belt (see Figure 4.6) with its screen facing

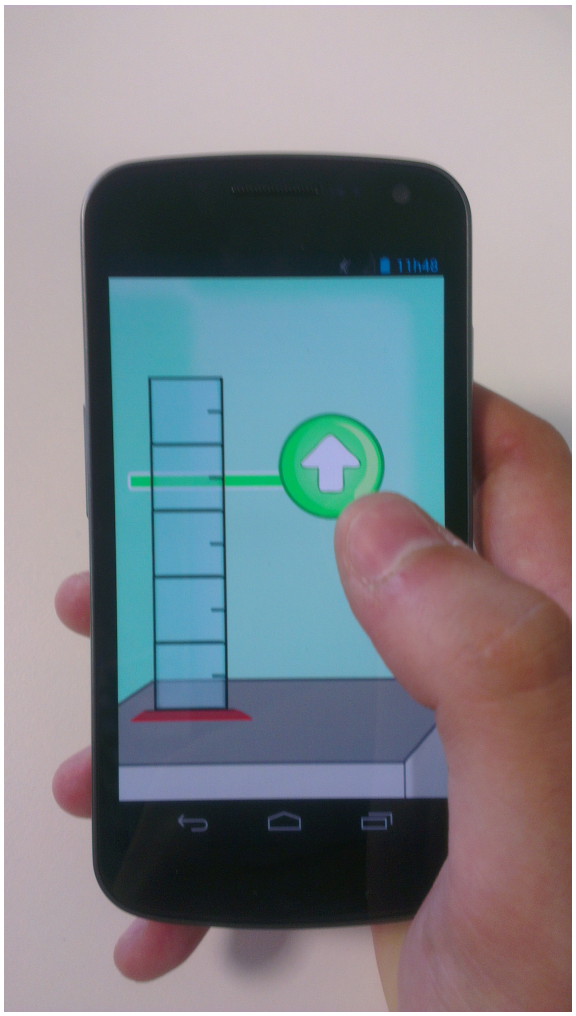

Figure 4.5: Layout of the second game incorporated into the project. Also it shows how the test is performed.

backwards and its top to the right. To be consistent with the International Society of Biomechanics (ISB) recommendations (Wu & Cavanagh, 1995) (see Figure 4.7) the signals suffered a direction transformation.

For the raw data to be readable it is processed through different filters to clean noise and errors and peak detection algorithms to detect the user readable data (Guimarães, 2011).

Since some data is necessary to start these series of algorithms (height, weight, leg size, foot length and shoe number) it is also a good opportunity to gather demographic information about the user (gender, age and, if PD patient, age when the disease was diagnosed). Aside from this data, it is possible to obtain the following features as well:

- **Number of steps** – From this number the first and last two steps are discarded and the rest are the ones analysed.
- **Total duration(seconds)** – Total walking time ($\frac{\text{number of foot contacts}}{\text{mean step length}}$).

Figure 4.6: Lower trunk position of an inertial sensor (Mancini & Horak, 2010).

- **Walking distance (meters)** – Distance that the user walked in the duration of the test (Nr. foot contacts $\times$ mean step length).
- **First leg in contact** – Represents the leg (left or right) with which the user started the test.
- **Temporal Measures:**
  - **Mean (right/left/both) steps duration (seconds)** – Mean of the time taken to take a step. This component differentiates left from right steps and calculates the total mean.
  - **Mean stride duration (seconds)** – Mean of the total time taken by the user to do a complete step (Mean right step duration + Mean left step duration).
  - **Mean stance phase (%)** – Percentage of the time where the user has the foot touching the ground (from the perspective of one foot, see Figure 4.8) ($\frac{\text{mean stance time}}{\text{mean stride time}}$).
  - **Mean swing phase (%)** – Percentage of the time where the user has the foot above the ground doing the step (from the perspective of one foot, see Figure 4.8) ($\frac{\text{mean swing time}}{\text{mean stride time}}$).
  - **Mean double support phase (%)** – Percentage of the time where the user has both feet touching the ground ($\frac{\text{mean double support time}}{\text{mean stride time}}$).
  - **Cadence (steps per minute)** – Estimative of total steps per minute the user does ($\frac{\text{Number of contacts}}{\text{total duration}}$).

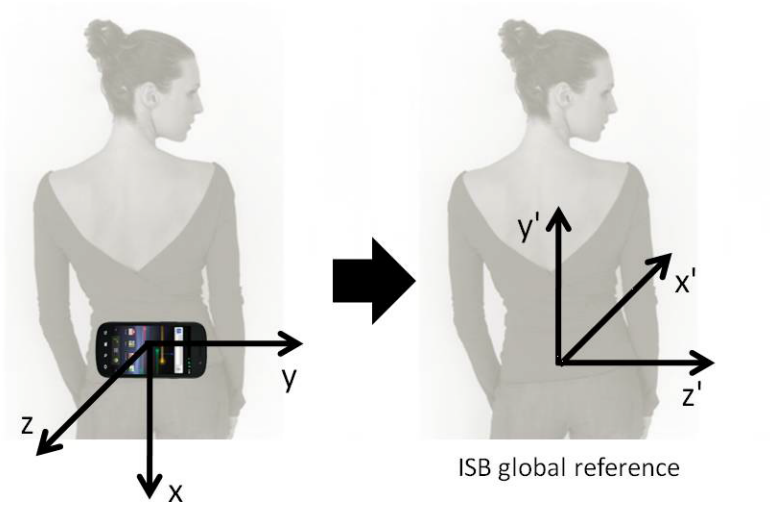

Figure 4.7: Phone coordinate system and ISB conventions in the reporting of kinematic data (Wu & Cavanagh, 1995).

- **Spatial Measures:**
  - **Mean (right/left/both) step length (meters)** – Mean of the distance covered by each step. This component, like in the temporal measures, differentiate right from left side.
  - **Walking speed (meters per second)** – Estimative of the walking speed of the user $(\frac{\text{walking distance}}{\text{total duration walking}})$.
- **Asymmetry:**
  - **R/L step duration asymmetry (%)** – Percentage of asymmetry in the step duration where the perfect value (no asymmetry) is zero $(\frac{(\text{mean right steps} - \text{mean left steps duration})}{\text{mean steps duration}})$.
  - **R/L step length asymmetry (%)** – Percentage of asymmetry in the step length $(\frac{(\text{mean right steps length} - \text{mean left steps duration})}{\text{mean steps length}})$.
- **Variability:**
  - **Stride time variability (standard deviation, milliseconds)** – All the values related to variability can help describe the walk the user did. The lower the more constant it was.
  - **Step time variability (standard deviation, milliseconds)**.
  - **Step length variability (standard deviation, millimetres)**.
- **Lateral Stability:**
  - **Pelvic sway (meters per** second^2 **-** m/s^2**)** – mean acceleration range on the transversal direction between steps (hip balance).

– **Lateral displacement (centimetres)** – mean lateral displacement reached on each gait cycle. It is an integration of the pelvic sway.

– **Lateral peak velocity (centimetre per second)** – maximum velocity reached on each cycle.

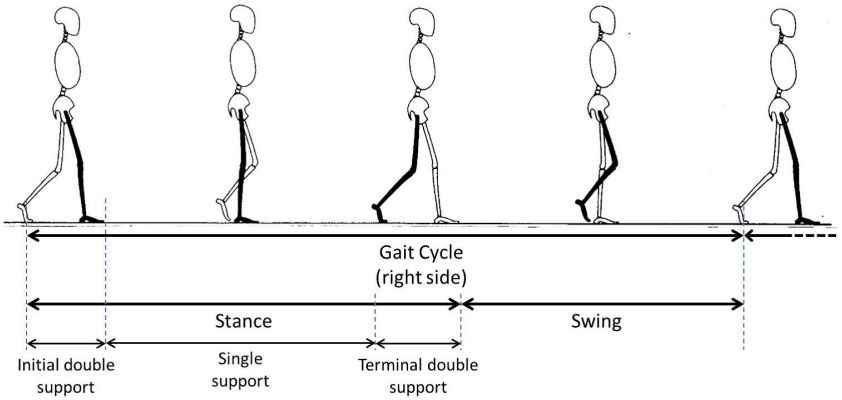

Figure 4.8: Gait cycle (Guimarães, 2011).

4.2 Data Format

All the data gathered will be stored in two types of files: Comma Separated Values (CSV) and Attribute-Relation File Format (ARFF). This type of files are used by different data mining applications (WEKA (Hall *et al.*, 2009) and RapidMiner (Mierswa *et al.*, 2006)). These applications allow to test different algorithms and compare them easily. Apart from the data files, log files from each component are created for backup purposes. Each log file stores information of each interaction of the user with the smartphone (click, coordinates, pressure, etc.).

4.3 Data Gathering Process

While waiting for the approval from the Ethics Commission of S.João's Hospital to start the gathering of data from PD patients the control group started to be assembled. When a considerable group had already performed the tests (18 subjects in a two week time line), a formal answer from the Hospital was received and authorization to start confirmed. After having 7 test subjects with PD it was easy to verify that the control group had an average age of 15 years over the PD patients group and another problem was that there were significant differences in the gait test between both groups, not because they were relevant but because there were differences in how the test was made since in the social centres were the control group members attend there could be some obstacles (other people, chairs, not enough length to perform the test always in the same direction). Therefore it was necessary to redo the control group inviting some younger subjects to

the Fraunhofer installations where it is possible to perform the test just as it is performed in the hospital. During the tests it was possible to conclude about the usability of the application. In some cases the spiral analysis component is impossible to be performed due to secondary effects of the medication that can cause muscular spasms. The same with the tapping games where the user can not pick up smartphone in the correct position. The test normally goes according to the following order: (1) Spiral, (2) Tap Game (right then left hand), (3) Water Tap Game (right then left hand), Gait and Simple Questions.

4.4 Decision Algorithm

Like explained in Section 3.2, different algorithms are to be tested in this problem and in some of them a FS algorithm must be used before. For the FS part, Dr. João Massano mentioned that information used for the asymmetry symptom must be relations between both hands and that all features regarding each hand individually should be discarded since the individual performance of each hand has no diagnosis information (regarding the tapping games only). Also personal information like age, height and weight is also discarded since it does not give any diagnosis information. In the gait test component there are also a few features that represent only one side of the body and features like the number of steps, distance covered and time taken with the test (which is fixed) that are too to be discarded. So for this phase a manual Filter Feature Selection was used and the final result was a subset of the original set of features.

Having the dataset with the non relevant features discarded (at least the ones that did not have any diagnose objective) the algorithm testing could be started. For that the RapidMiner was used (incorporating all the WEKA algorithms as a plug-in) since it provides a very friendly user interface (drag and drop basis). Due to its simplicity in the output as well it is possible to compare the algorithms easily. Since one of the main objectives of the project is to find features that can indicate traces of PD, using different ranking and FS algorithms it is possible to verify which ones are the most relevant. However several classification algorithms (like trees) also determine the most relevant ones (the first selected is the most).

Chapter 5

Results

The result analysis are separated in four parts: the spiral component, data analysis, feature selection analysis and algorithm comparison. In the spiral component it is demonstrated the difference between drawing the spiral with and without the stylus pen. In the second one simple statistical measures are calculated to better understand the features distribution between PD patients and the control group. In the third one different types of features selection and feature ranking algorithms are used on the features selected after removing the ones discussed with Dr. Massano and the ones with no obvious diagnosing relevance. The final part compares the machine learning algorithms described in Section 3.3 using the methods mentioned in Section 3.2. In the end all the result are discussed together explaining the new findings and what has been proven or not.

5.1 Spiral

Like mentioned in the Section 2.1.2.1, this component was originally developed for a tablet using a common stylus. In this case the use of a smartphone could limit the potentiality of this component. So a new type of stylus was used (Adonit, 2013), that in theory would facilitate the drawing of the spiral since it has a transparent end. To validate this solution a small test group (five healthy persons) was gathered to perform the test with their index finger. Their results are compared to the rest of the control group gathered for the main problem. One of the main complaints this small group had while performing the test was the lack of vision available on the screen while drawing the spiral due to the size of their finger. This complaint was never mentioned by the control group (nor the PD patients), that used the stylus pen (see Figure 5.1). It is to remember that a regular stylus replicates the index finger's end and have similar diameter so it is likely possible that the same complaint would occur in case of using it. In the Table 5.1, it is possible to see the result obtained in this experiment where the control group had a 9.7 ± 4.66 pixels average error with the stylus and the small group without it had a 17.3 ± 4.33 pixels of average error. The Figure 5.2, represent an example of each case (with and without stylus).

	Finger	Stylus
Average Error (px)	17.32	9.94
Standard Deviation (px)	4.32	4.14

Table 5.1: Comparison between drawing the spiral with the finger or the stylus

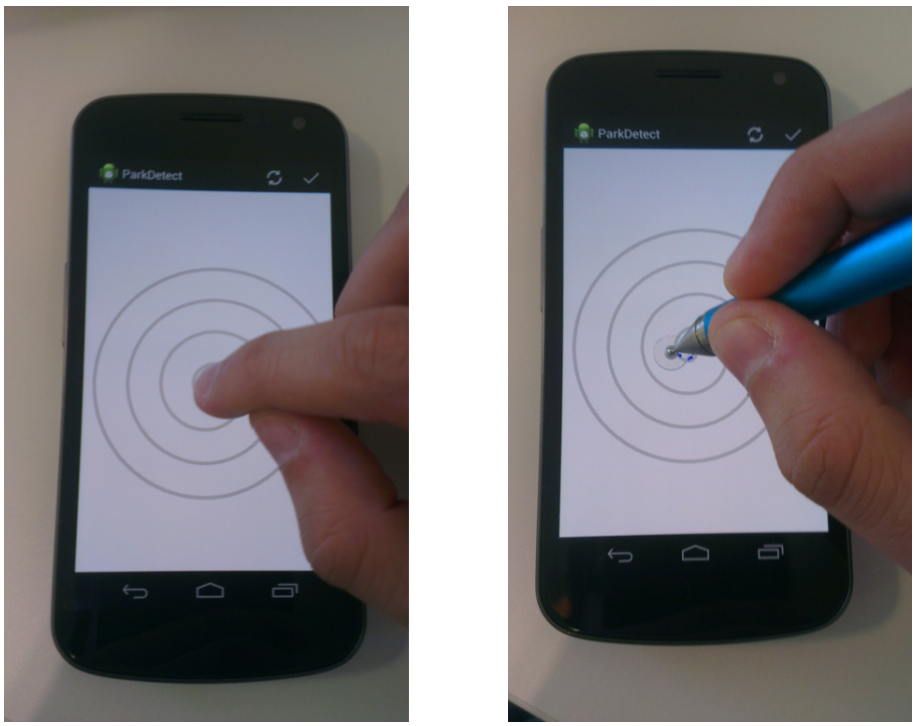

Figure 5.1: Example of how the user draws the spiral with the finger and with the stylus

5.2 Data

After discarding the features that posed no relevance (from one side only and height, weight, etc.) there are still plenty of features to be analysed and considered. The average age of PD patients was 67.5 ± 3.97, 12 male and 5 female, and the control group of 72.1 ± 9.03, 14 male and 4 female. On average the PD patients have been diagnosed for 8 ± 6.6 years (however the older cases did not show any extreme symptoms that could affect the whole group). In the Table 5.2, there are all the features and their statistical values in order to easily compare them. Also the Figure 5.3 demonstrates an example with a Spiral drawn by a PD patient and an healthy person from the control group.

5.3 Feature Selection

Like previously stated in Section 3.1 different algorithms can be used to try to verify the relevance of the features gathered for classification purposes. Keeping that in mind this part is only for curiosity purposes since the algorithms have internal methods to deal with irrelevant features.

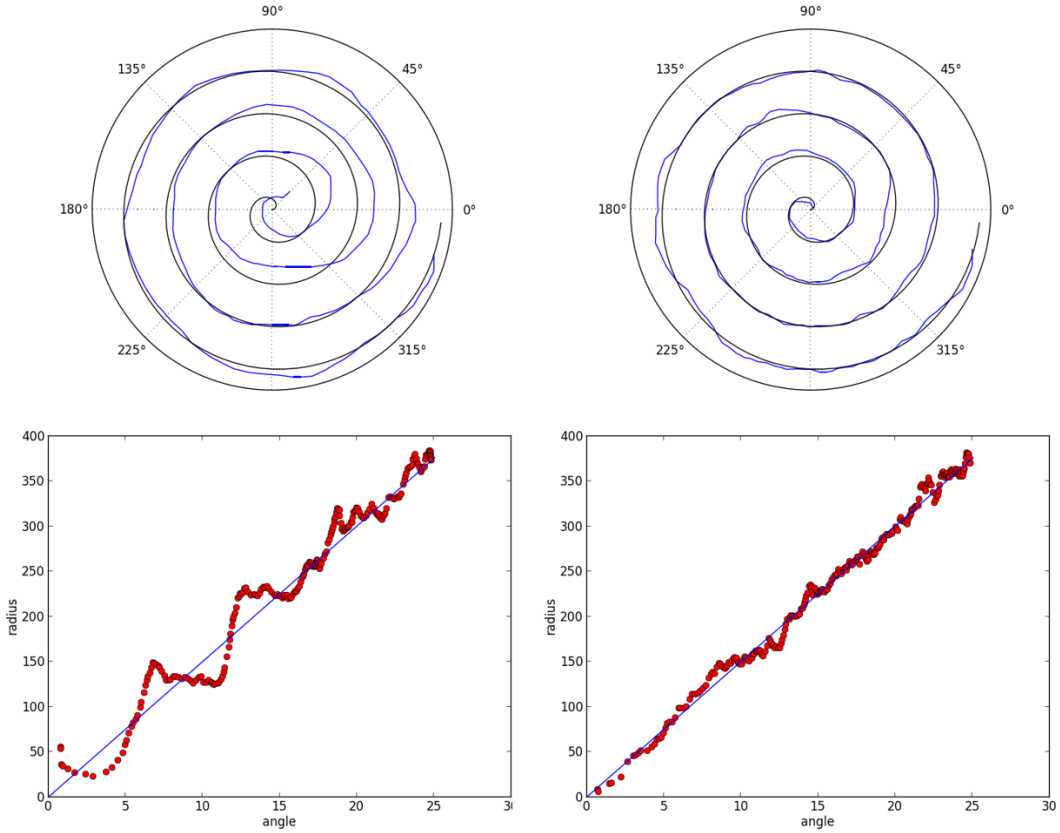

Figure 5.2: Representative example of a spiral drawn with the finger (left side) and with the stylus (right side). The upper images are in Polar Coordinates and the lower in Cartesian Coordinates. Both of them have the perfect spiral coordinates represented.

Using the Information Gain Attribute Evaluation method, that verifies which features give more information for classification purposes, there are 6 features that pop up with some relevance value (zero is non relevant, one most relevant):

- Pelvic Sway – 1.0
- Rest Tremor – 0.613
- Walking Speed – 0.580
- Spiral Cross – 0.476
- Posture – 0.388
- Tap Time Ratio – 0.387

However there are some algorithms that just verify which subset of features have the most relevance together, not single features. Using a Best First (greedy hill climbing) algorithm it selects the exact same features given in the previous test, but using a Greedy Stepwise (SBE) one,

Feature	PD	Control Group
Spiral Average Error (px)	21.66 ± 23.94	9.94 ± 4.14
Spiral Cross (%)	6.6 ± 3.7	11 ± 2.9
Spiral Pressure Ratio (%)	3.1 ± 3	3.3 ± 3.1
Spiral Side Ratio (%)	13.5 ± 10.9	16.3 ± 10.4
Tap Time Ratio (%)	33.5 ± 30.8	13.4 ± 9.9
Tap Down Time Ratio (%)	22.5 ± 14.3	35.9 ± 15.4
Tap Pressure Ratio (%)	11.7 ± 8	13 ± 9.6
Water Time Ratio (%)	28 ± 19.7	10.4 ± 10.8
Water Pressure Ratio (%)	10 ± 14	6.6 ± 9.6
Water Speed Ratio (%)	27.8 ± 25.8	9.3 ± 8.1
Flexed Posture (% Positive)	29.4	83.3
Rest Tremor (% Positive)	11.8	77.8
Mean Steps Duration (Seconds)	0.58 ± 0.17	0.477 ± 0.072
Mean Stride Duration (Seconds)	1.16 ± 0.35	0.95 ± 0.14
Mean Stance Phase (%)	57.41 ± 2.09	58.88 ± 2.45
Mean Swing Phase (%)	42.54 ± 1.9	41.18 ± 2.35
Mean Double Support Phase (%)	7.44 ± 1.97	8.86 ± 2.43
RL Duration Asymmetry (%)	8.33 ± 9.17	4.87 ± 4.07
Stride Time Variability (Milliseconds)	100 ± 126.36	41.32 ± 22.92
Step Time Variability (Milliseconds)	79.45 ± 71.57	38.41 ± 19.06
Cadence (steps per Minute)	108.75 ± 19.23	130.6 ± 31.92
Walking Speed (Meters per Second)	1.25 ± 0.265	1.64 ± 0.51
Mean Step Length (Meters)	0.688 ± 0.071	0.75 ± 0.079
Step Length Variability (Centimetres)	32.94 ± 16.14	29.68 ± 8.14
RL Length Asymmetry (%)	3.59 ± 3.03	2.36 ± 1.76
Pelvic Sway (Meters per Second^2)	3.24 ± 0.64	5.64 ± 1.63
Lateral Displacement (Centimetres)	3.01 ± 1.6	2.12 ± 0.7
Lateral Peak Velocity (Centimetre per Second)	15.03 ± 4	15.61 ± 3.48

Table 5.2: All features gathered where the values are average $\pm$ Standard Deviation

it only select 4 of them (Tap Time Ratio, Posture, Rest Tremor and Pelvic Sway). All this results were obtained using Attribute Selection algorithms implemented in the Weka package using the dataset with the features mentioned in Section 4.4 discarded.

5.4 Classification Algorithms

Using the Rapid Miner graphical interface it is possible to test different algorithms and evaluate their performance. All the algorithms were tested using 10-Fold Cross Validation. The three algorithms used are: Decision Trees (C4.5, since C5.0 can only be used online and the most of the improvements made from it ancestor are scalability issues), RipperK rules and Bayesian Networks. For each the values of the Accuracy, Precision, Recall and AUC are calculated (see Table 5.3) and the Confusion Matrix (see Tables 5.4, 5.5 and 5.6) and ROC curves shown (see Figures 5.4, 5.5 and 5.6).

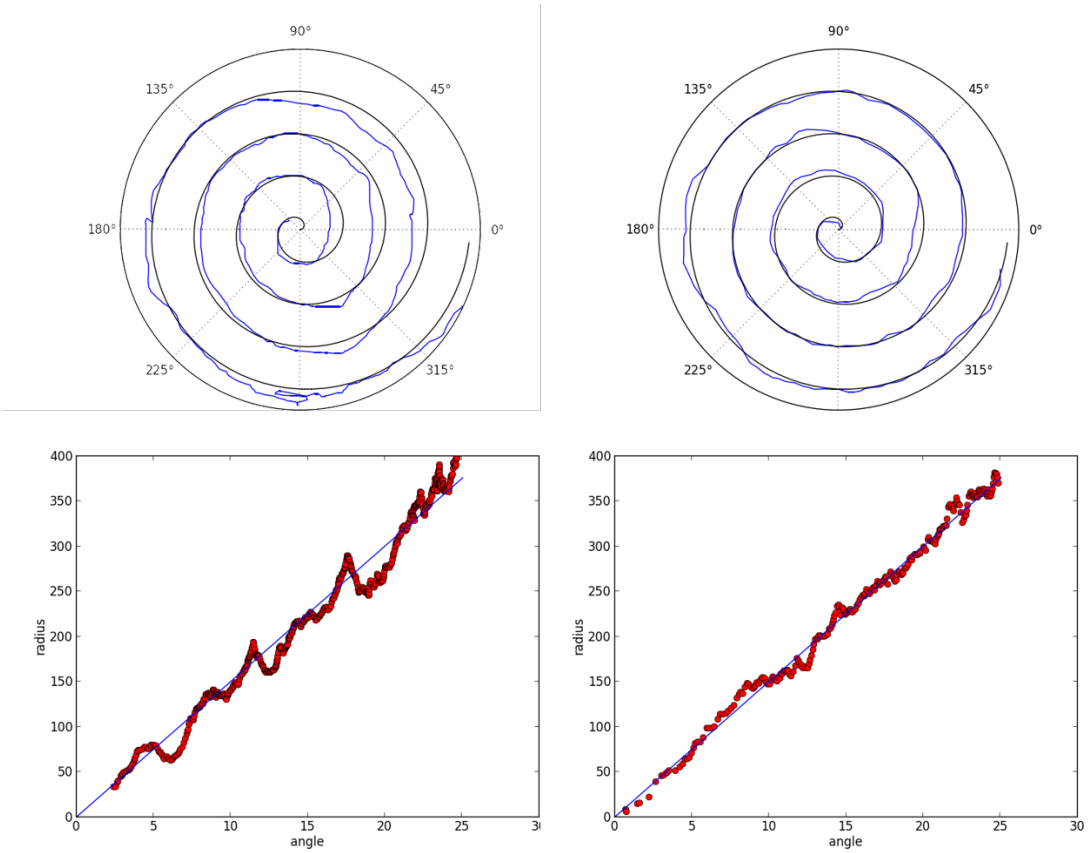

Figure 5.3: Representative example of a spiral drawn by a PD patient (left side) and by an healthy individual (right side). The upper images are in Polar Coordinates and the lower in Cartesian Coordinates. Both of them have the perfect spiral coordinates represented.

Measure	**C4.5**	**RipperK**	**Bayesian Networks**
Accuracy (%)	86.67 ± 13.54	80.83 ± 17.10	87.5 ± 23.05
Precision (%)	91.67 ± 17.08	80.83 ± 20.43	86.67 ± 30.55
Recall (%)	86.67 ± 20.82	90 ± 20	85 ± 32.02
AUC	0.825 ± 0.195	0.475 ± 0.261	0.875 ± 0.202

Table 5.3: Comparison between the algorithms tested. Since a 10-cross fold validation was used the values presented are the average $\pm$ standard deviation.

	true Negative	true Positive	class precision
predicted Negative	15	3	83.33%
predicted Positive	2	15	88.24%
class recall	88.24%	83.33%	

Table 5.4: Confusion Matrix for the C4.5 Decision Tree algorithm

Also the models are exportable where the C4.5 and the RipperK are easily readable (see Figures 5.7 and 5.8), but the Bayesian Network is not (only when classifying a new instance is possible

	true Negative	true Positive	class precision
predicted Negative	12	2	85.71%
predicted Positive	5	16	76.19%
class recall	70.59%	88.89%	

Table 5.5: Confusion matrix for the RipperK algorithm

	true Negative	true Positive	class precision
predicted Negative	15	3	83.33%
predicted Positive	2	15	88.24%
class recall	88.24%	83.33%	

Table 5.6: Confusion matrix for the Bayesian Network algorithm

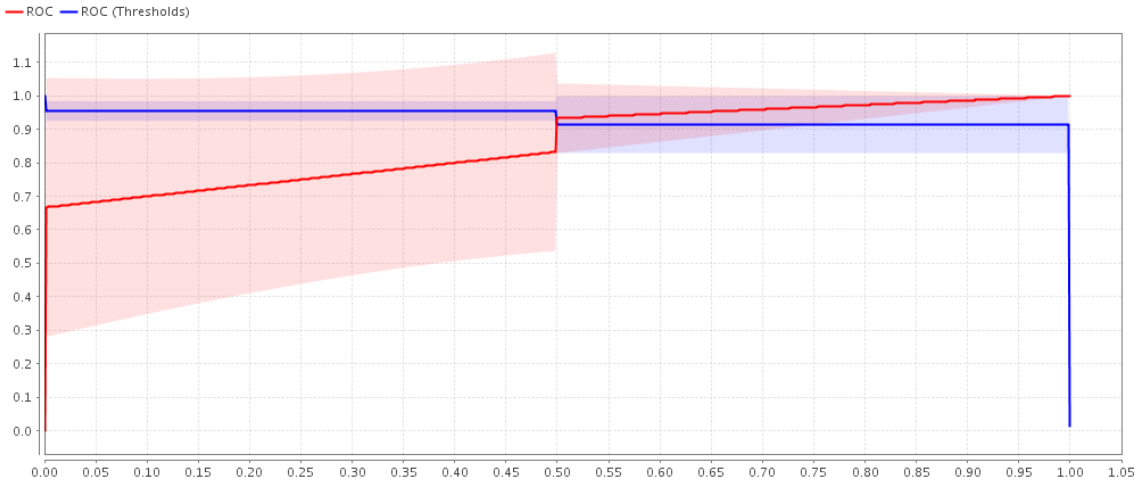

Figure 5.4: ROC Curve for the C4.5 Decision Tree algorithm

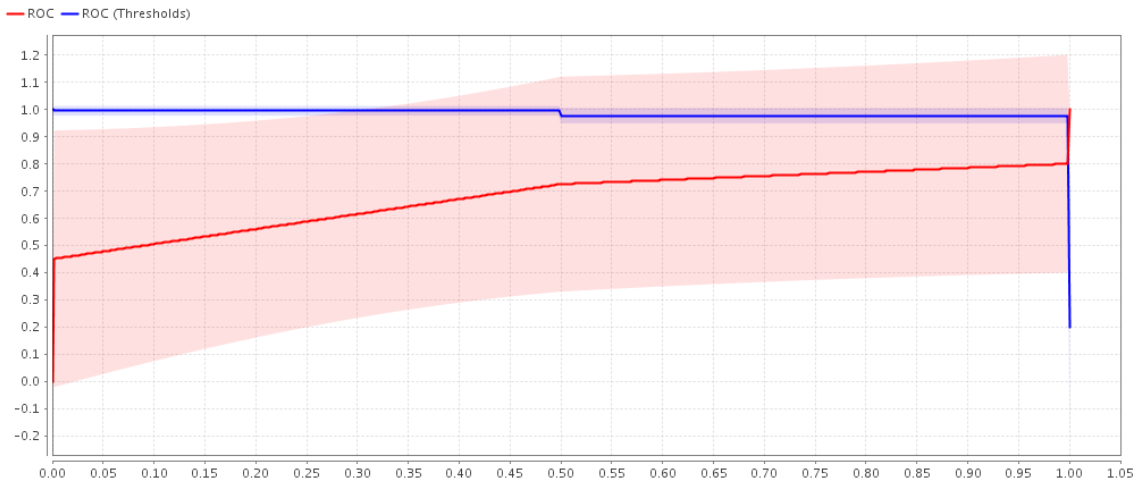

Figure 5.5: ROC Curve for the RipperK algorithm

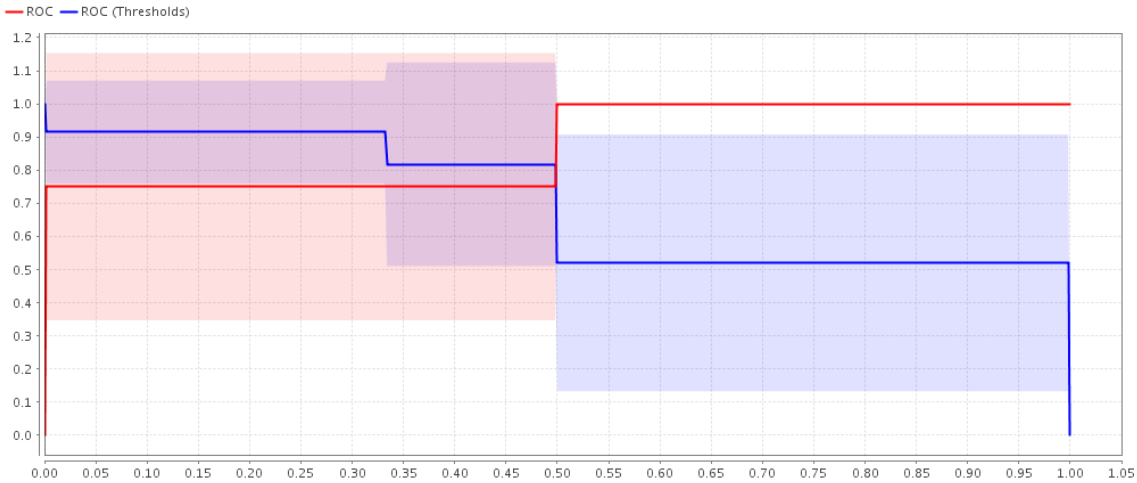

Figure 5.6: ROC Curve for the Bayesian Network algorithm

to see the reasoning behind the decision made). There model are easily exported to java through the use of the weka package or just translating the obtained model to Java by hand.

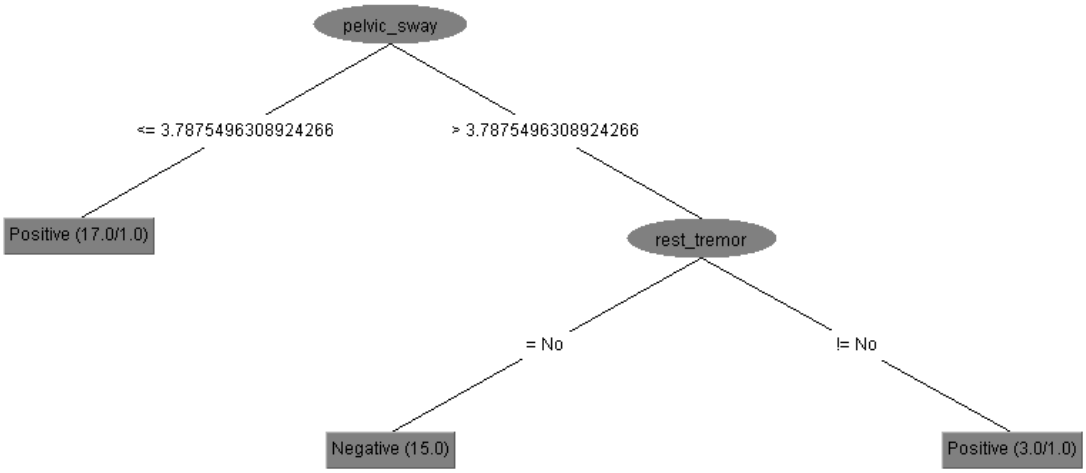

Figure 5.7: Model obtained from the Decision Tree's algorithm using the whole dataset for training since it is impossible to obtain a model from the 10-fold cross validation.

5.5 Discussion

From the start it was known that one of the main step backs for this project was the amount of data that would be available at the end. Due to a delay from S. João's Hospital and even when the tests started most of the PD patients that have regular appointments are not capable of performing them or are in very late stages of the disease it was not possible to have a relevant number of subjects to state that the results obtained have a high degree of confidence. However they do seem to have

```
JRIP rules:
===========

(pelvic_sway >= 3.928434) and (spiral_mean_error <= 11.436671) => PD=Negative (15.0/0.0)
(spiral_cross >= 0.09375) and (posture = No) => PD=Negative (2.0/0.0)
 => PD=Positive (18.0/0.0)

Number of Rules : 3
```

Figure 5.8: Model obtained from the RipperK algorithm using the whole dataset for training since it is impossible to obtain a model from the 10-fold cross validation.

promising future if the project continues with the data gathering process. Also some modifications to the control group had to be made to maintain a demographic consistency between both groups which delayed the algorithm testing phase.

Following the results structure, the idea of getting a different kind of stylus for the spiral component played off since the subjects that performed the test with the finger had almost two times the average error from the ones with the stylus (with roughly the same variability). Normally using the finger the user ca not really know where he is pressing and it results on a spiral shifted to one side on the parts the user loses complete vision of the path (right side of the spiral when right handed). However even with the stylus some PD subjects were unable to draw the spiral due to muscular stiffness or spasms on the hand or arm.

From the data considered after the meeting with Dr. Massano, some promising results were obtained across the four components of the application. In the spiral component some of the PD patients had a certain level of difficulty drawing it that justifies the high variability in all its features. However the difference between the two groups is quite evident with the PD group having over twice the average pixels or error. These results were similar to the ones obtained in (Surangsrirat & Thanawattano, 2012) which does not discard the use of a device with a smaller screen to obtain relevant data. Analysing the Figure 5.3, even if both of the spirals have similar pixel errors, the control group one have a constant and linear error, while on the PD one the error is not constant and with peaks on opposite directions in short segments of the spiral. Also PD patients have a tendency to lift the stylus from the screen more times than the control group.

Across almost all features it is visible that the data from the PD group has a high variability which means that among this group some subjects performed the test quite well in terms of obvious symptoms. In the tap component there is a visible difference in the time ratio from one group to another but the two other features have no relevant difference since when the subject from the control group performed the test they had a constant high speed and kept the finger on the screen for a very small amount of time which in a milliseconds scale caused a high difference between left and right hands. The difference in the water component is more evident. All three features have a substantial gap, even considering the variability which means that the tapping using the thumb can detect an asymmetry in the hands easier than the regular tapping with the index finger.

The features regarding the flexed posture and the rest tremor have the expected results since

the symptoms are common on PD patients and rare in the control group. Although the control group was gathered taken into consideration possible difficulties by their part on performing the test. Therefore some subjects of the control group have some symptoms similar to the ones of PD but, in this case, not related to the disease.

The final component is where the most promising results are. Because it is a passive test where it requires no concentration other than walking non-stop in a straight line avoiding possible obstacles the data is the most reliable. In theory, because of the slowness of movements caused by PD, it was expected from PD patients that they had smaller steps in distance, longer steps in time and, therefore, lower velocity. That was confirmed in this test where all features related with step length, time and velocity showed differences between both groups. The asymmetry related features also show some differences where higher values are associated to the PD group. But the feature that has the most difference between both groups is the Pelvic Sway that not only has a great difference in its average but also have a low variability. These values are also related to the slowness of movements that make the PD subjects balance their hips less than what a regular person would.

Knowing the difference across the features from both groups it is possible to say which ones could have more relevance. However they might not be so relevant for classification purposes. Using the algorithms mentioned only 6 features showed classification relevance. All these features also had differences in the statistical analysis of the data.

The algorithm testing phase was affected with the lack of subjects available that justifies the use of the 10-fold cross validation. Still the final results are very promising but not inspiring. Both C4.5 and RipperK mainly base their decision on the pelvic sway feature (just a small difference on the test parameter). This is a result of the lack of data that after having not great confidence in continuing to perform more tests the algorithm ends there. If there were more cases of that type the algorithm would have more confidence and probably test with some other relevant feature. Still looking at the measures obtained the three algorithms obtained over 80% accuracy, with the Bayesian Networks with the highest value but also the highest variability across the 10 folds. In terms of precision all above 80 % as well but in this case the best one is the C4.5 regarding average and variability. The recall is where the RipperK shines with 90% but has low AUC on opposite for its counterparts. Looking at the confusion matrixes the C4.5 and the Bayesian Networks have identical results missing 5 out of 35 classifications. The RipperK may have missclassified 7 out of 35 but most of them are false-positive and has a lower number of false-negatives which in this case might have higher importance. Comparing the ROC curves it is possible to see that the RipperK has a lower performance and the C4.5 and the Bayesian Networks similar performance.

Like previously stated, the lack of data also affects the decision on which algorithm to use and a final decision cannot be made without at least twice the amount of data available (at least 75 subjects) to reach properly justified conclusions. Still with this small group it was possible to verify the importance of an application capable of detecting some symptoms of PD to accelerate the process of sending the patient to a neurologist because the sooner the patient gets medicated, the better quality of life he is going to get. Although the results obtained cannot be used to develop

an application highly capable of detecting PD, it was possible to validate different theories that were mentioned throughout the development of this project.

Chapter 6

Conclusions and Future Work

After several internal meetings and with Dr. Massano, the different components of the application were defined and implemented. However there was a delay from S.João's Hospital. Meanwhile the control group started to be gathered. When the tests started with Parkinson's Disease patients it was possible to see a gap in the average age of both groups (with the control group much older), so modifications to the control group had to be done as well. With only one afternoon per week in the Hospital the whole project had a step back. However with 35 subjects (both groups) it was possible to start testing different machine learning algorithms that showed promising results overall, but still not sufficient to justify the use of a particular one.

6.1 Objectives Fulfilment

From the start the main idea was to detect different symptoms of Parkinson's Disease through the use of the components existent in a smartphone only. From Dr. Massano's point of view if at least one symptom could be properly detected it would be a great accomplishment. Taking into consideration the lack of data all affirmations do not have a high confidence rating but it is possible to say that the application can detect patterns of hand asymmetry, slowness of movements and gait velocity and, so far, the values captured have substantial differences from healthy persons.

6.2 Future Work

For future work, the main objective would be gathering more data from new test subjects that can only be countered with more available time. After the whole process of data gathering is was possible to verify that a modification to the tapping games would greatly increase its ability to detect not only hand asymmetry but also the slowness of movements if the game is modified to increase to number of repetitive taps in each loop to at least 10. That way it becomes a more consistent source of data and it is easier to verify if there is a slowness in the later taps. Also a

rework on the application's interface in a way that any doctor can understand it and tutorials for each component so all the subjects perform the test the correct way. Another good investment would be the centralization of all information in a server, where there would be two kinds of interaction: a message to get a classification on a subject or add a new subject to the data set to enlarge it even more. Of course the second option would only be possible to be used by specialists but that way the server could update the classification model in real time, and all applications would use the same model without any need of an update.

References

Adonit. 2013. *Jot Classic Fine Point Stylus for iPad and Android*. available at `www.adonit.net/jot/classic/`, consulted on 18/02/2013.

Allen, R. G., & Tresini, M. 2000. Oxidative stress and gene regulation. *Free Radic. Biol. Med. 28*, 680–683.

Aly, N. M., Playfer, J. R., Smith, S. L., & Halliday, D. M. 2007. A novel computer-based technique for the assessment of tremor in Parkinson's disease. *Age and Ageing, vol. 36*, 395–399.

Baken, R.J., & Orlikoff, R.F. 2000. *Clinical measurement of speech and voice*. Second edn. San Diego: Singular Thomson Learning.

Bowers-Gentry, R. C., Deems, R. A., Harkewicz, R., & Dennis, E. A. 2006. Eicosanoid lipidomics. *Functional Lipidomics. CRC Press-Taylor and Francis FGroup, Boca Ranton*, 79–100.

Cattlet, Jason. 1991. Megainduction: a test flight. *In Proc. of the 8th Int. Workshop on Machine Learning, Ithaca, New York.*

Charniak, Eugene. 1991. Bayesian Networks without Tears. *AI Magazine Volume 12 Number 4*.

Cohen, William W. 1995. Fast Effective Rule Induction. *Machine Learning: Proc. of the 12th International Conference*.

Cunnungham, L., Nugent, C., Moore, G., Finlay, D., & Craig, D. 2009. Computer-based assessment of bradykinesia, akinesia and rigidity in Parkinson's disease. *Proceedings of ICOST*.

Fahn, S., & Elton, R.L. 1986. Unified Parkinson's disease rating scale. *Fahn S, Marsden CD, Goldstein M, et al, eds. Recent developments in Parkinson's disease. Vol II.*, January, 153–163.

Farooqui, Tahira, & Farooqui, Akhlaq A. 2009. Aging: An important factor for the pathogenesis of neurodegenerative diseases. *Mechanisms of Ageing and Development 130*, 203–215.

Fürnkranz, Johannes, & Widmer, Gerhard. 1994. Incremental reduced error pruning. *Machine Learning: Proc. of the 11th Annual Conference*.

Gandhewar, N., & Sheikh, R. 2011. Google Android: an emerging software platform for mobile devices. *International Journal on Computer Science and Engineering*, February, 12—17.

German, J. B., Gillies, L. A., Smilowitz, J. T., Zivkovic, A. M., & Watkins, S. M. 2007. Lipidomics and lipd profiling in metabolomics. *Curr. Opin. Lipidol. 18*, 66–71.

Guimarães, Vânia. 2011. *Phone Based Fall Rist Detection*. M.Phil. thesis, Faculdade Engenharia Universidade do Porto.

REFERENCES

Hall, Mark, Frank, Eibe, Holmes, Geoffrey, Pfahringer, Bernhard, Reutemann, Peter, & Witten, Ian H. 2009. *The WEKA Data Mining Software: An Update*.

Harel, B., Cannizzaro, M., & Snyder, P.J. 2004. Variability in fundamental frequency during speech in prodromal and incipient Parkinson's disease: A longitudinal case study. *Brain and Cognition 56*, 24—-29.

Hastie, Trevor, Tibshirani, Robert, & Friedman, Jerome. 2001. *The Elements of Statistical Learning: Data Mining, Inference and Prediction*. Springer.

Ho, A., Iansek, R., Marigliani, C., Bradshaw, J., & Gates, S. 1998. Speech impairment in a large sample of patients with Parkinson's disease. *Behavioral Neurology 11*, 131–37.

Hopley, Lara, & van Schalkwyk, Jo. 2001 (September). *The magnificent ROC*. available at `http://www.anaesthetist.com/mnm/stats/roc/Findex.htm`, consulted on 20/05/2013.

Jankovic, Joseph. 2008. Parkinson's disease: clinical features and diagnosis. *Journal of Neurology, Neurosurgery and Psychiatry with Practical Neurology*, **1**, 368–376.

John, G. H., Kohaviand, R., & Pfleger, K. 1994. Irrelevant Features and the Subset Selection Problem. *In Proc. of the 11th Int. Conf. on Machine Learning*, 121—129.

Little, M.A., McSharry, P.E., Hunter, E.J., Spielman, J., & Ramig, L.O. 2009. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Biomedical Engineering 56(4)*, 1015–1022.

Liu, X., Carrol, C. B., Wang, S. Y., Zajicek, J., & Bain, P. G. 2005. Quantifying drug-induced dyskinesias in the arms using digitised spiral-drawing tasks. *Journal of Neuroscience Methods, vol 144*, 47–52.

Lu, Y., Hong, S., Gotlinger, K., & Serhan, C. N. 2006. Lipic mediator informatics and proteomics in inflammation-resolution. *The Scientific World J. 6*, 589–614.

Mancini, M., & Horak, F. B. 2010. The relevance of clinical balance assessment tools to differentiate balance deficits. *European journal of physical and rehabilitation medicine, 46(2)*, 239–248.

Massano, João, & Bhatia, Kailash P. 2012. Clinical Approach to Parkinson's Disease: Features, Diagnosis, and Principles of Management. *Cold Spring Harb Perspect Med 2012*.

Mierswa, Ingo, Wurst, Michael, Klinkenberg, Ralf, Scholz, Martin, & Euler, Timm. 2006. YALE: Rapid Prototyping for Complex Data Mining Tasks. *Pages 935–940 of:* Ungar, Lyle, Craven, Mark, Gunopulos, Dimitrios, & Eliassi-Rad, Tina (eds), *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. New York, NY, USA: ACM.

Miralles, F., Tarongi, S., & Espino, A. 2006. Quantification of the drawing of an Archimedes spiral through the analysis of its digitized picture. *Journal of Neuroscience Methods, vol. 152*, 47–52.

Mitchell, T. M. 1982. Generalization as Search. *Artificial Intelligence, 18(2)*, 203—226.

REFERENCES

Molina, Luis Carlos, Belanche, Lluís, & Àngela Nebot. 2002. Feature Selection Algorithms: A Survey and Experimental Evaluation. *ICDM '02 Proceedings of the 2002 IEEE International Conference on Data Mining*, 306.

Nieto-Sampedro, M., & Nieto-Diaz, M. 2005. Neural plasticity: changes with age. *J. Neural Tansm. 112*, 3–27.

Nisbet, Robert, IV, John Elder, & Miner, Gary. 2009. *Handbook of Statistical Analysis and Data Mining Applications*. Academic Press, Elsevier.

Pagallo, Giulia, , & Haussler, David. 1990. Boolean feature discovery in empirical learning. *Machine Learning 5 (1)*.

Pearl, Judea. 2003. *Bayesian Networks, Causal Inference and Knowledge Discovery*. available at `http://www.secondmoment.org/articles/bayesian.php`, consulted on 22/05/2013.

Peters, A. 2002. The effects of normal aging on myelin and nerve fibers: a review. *J. Neurocytol. 31*, 581–593.

Pullman, S. L. 1998. Spiral analysis: a new technique for measuring tremor with a digitizing tablet. *Movement Disorders vol. 13*, 85–89.

Pullman, S. L., Derby, C., Floyd, A., Bressman, S., & Lipton, R. B. 2008. Validity of spiral analysis in early Parkinson's disease. *Movement Disorders, vol 23*, 831–537.

Qin, L., Liu, Y., Cooper, C., Liu, B., Wilson, B., & Hong, J. S. 2002. Microglia enhances B-amyloid peptide-induced toxicity in cortical and mesencephalic neurons by producing reactive oxygen species. *J. Neurochem. 83*, 973–983.

Quinlan, Ross. 1987. Simplifying decision trees. *International Journal of Man-Machine Studies, 27*, 221—234.

Quinlan, Ross. 2008 (October). *Ross Quinlan's personal's webpage*. available at `http://www.rulequest.com/Personal/,consultedon21/05/2013`.

RuleQuest. 2012 (February). *Information on See5/C5.0*. available at `http://rulequest.com/see5-info.html`, consulted on 21/05/2013.

Sapir, S., Ramig, L., Spielman, J., & Fox, C. 2010. Formant Centralization Ratio (FCR): A proposal for a new acoustic measure of dysarthric speech. *Journal of Speech Language and Hearing Research, 53*, 114–125.

Savitt, Joseph M., Dawson, Valina L., & Dawson, Ted M. 2006. Diagnosis and treatment of Parkinson disease: molecules to medicine. *The Journal of Clinical Investigation*, July, 1744–1754.

Schoentgen, J., & Gucteneere, R. De. 1995. Time series analysis of jitter. *Journal of Phonetics, Vol. 23*, 189–201.

Surangsrirat, Decho, & Thanawattano, Chusak. 2012. Android Application for Spiral Analysis in Parkinson's Disease. *2012 Proceedings of IEEE Southeastcon*, March, 1—6.

REFERENCES

TekEye. 2013. *What are DPI, DIP, DP, PPI, SP and Screen Resolutions in Android? |Tek Eye*. available at `http://tekeye.biz/2013/android-dpi-dip-dp-ppi-sp-and-screens`, consulted on 06/02/2013.

Titze, I.R. 2000. *Principles of Voice Production*. Second edn. National Center for Voice and Speech, Iowa City, US.

Tsanas, A., Little, M.A., McSharry, P.E., & Ramig, L.O. *Using the cellular mobile telephone network to remotely monitor Parkinson's disease symptom severity*.

Tsanas, A., Little, M.A., McSharry, P.E., & Ramig, L.O. 2010. Accurate telemonitoring of Parkinson's disease progression using non-invasive speech tests. *IEEE Trans. Biomedical Engineering 57*, 884–893.

Tsanas, A., Little, M.A., McSharry, P.E., & Ramig, L.O. 2011. Nonlinear speech analysis algorithms mapped to a standard metric achieve clinically useful quantification of average Parkinson's disease symptom severity. *Journal of the Royal Society Interface, Vol. 8*, 842–855.

Tsanas, A., Little, M.A., McSharry, P.E., Scanlon, B.K., & Papapetropoulos, S. 2012a. Statistical analysis and mapping of the Unified Parkinson's Disease Rating Scale to Hoehn and Yahr staging. *Parkinsonism and Related Disorders*, June, 697–9.

Tsanas, Athanasios, Little, Max A, Mcsharry, Patrick E, Spielman, Jennifer, & Ramig, Lorraine O. 2012b. Novel Speech Signal Processing Algorithms for High-accuracy Classification of Parkinson's Disease. *IEEE Trans. Biomedical Engineering 59*, 1264–1271.

Wang, H., Yu, Q., Kurtis, M. M., Floyd, A. G., Smith, W. A., & Pullman, S. L. 2008. Spiral analysis - Improved clinical utility with center detection. *Journal of Neuroscience Methods, vol. 171*, 264–270.

Wang, M., Wang, B., Zou, J., Zhang, J., & Nakamura, M. 2011. Quantitive evaluation of hand movement in spiral drawing for patients with parkinsons's disease based on modeling in polar coordinate system with varied origin. *Proceeding of IEEE/ICME on Complex Medical Engineering*, 169–173.

Watson, A. D. 2006. Lipidomics: a global approach to lipid analysis in biological systems. *J. Lipid Res. 47*, 2101–2111.

Weiss, Sholom, & Indurkhya, Nitin. 1991. Reduced complexity rule induction. *In: 12th Int. Joint Conference on Artificial Intelligence*.

Westin, J., Ghiamati, S., Memedi, M., Nyholm, D., Johansson, A., & Dougherty, M. 2010. A new computer method for assessing drawing impairment in Parkinson's disease. *Journal of Neuroscience Methods, vol. 190*, 143–148.

Wu, Ge, & Cavanagh, Peter R. 1995. ISB Recommendations in the Reporting for Standardization of Kinematic Data. *Journal of Biomechanics, 28(10)*, 1257–1261.

Yu, Q., Pullman, S. L., Fahn, S., & Pedersen, S.F. 1997. Homonymous hemispiral abnormalities in patients with Parkinson's disease. *Neurosci Abstr, vol. 23*, 1898.